DOCUMENT RESUME

ED 188 182 CS 205 604

AUTHOR Mersand, Joseph
TITLE In Quest of the Perfect Test.
PUB DATE 80
NOTE 16p.: Reprint from "High Points," October 1979.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Testing; *English Instruction; Essay Tests; Higher Education; Language Arts; Multiple Choice Tests; Secondary Education; *Test Construction; Test Items

ABSTRACT

This booklet offers English teachers a summary of the literature on test development and practical suggestions for test writing. The issues addressed include (1) honesty in testing, (2) understanding the place of testing, (3) characteristics of good testing, (4) preparing a classroom test, (5) what diagnostic tests reveal, (6) essay questions, and (7) objective questions (true-false, multiple choice, and matching). Included are samples of various types of test questions. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED188182

IN QUEST OF THE PERFECT TEST

By

Joseph Mersand, Ph.D.

Instructor

Institute of New Dimensions

Palm Beach Junior College

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joseph Mersand

Kenneth R. Fanizzi

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1980

CS205604

Reprinted by permission of
Kenneth R. Fanizzi
Editor, High Points
Board of Education of the
City of New York

Copyright, 1980 by
Joseph Mersand

IN QUEST OF THE PERFECT TEST

Joseph Mersand

In 1926, William Dana Orcutt, famous printer and novelist, wrote an illuminating book *In Quest of the Perfect Book,*[1] in which he related his exciting adventures for thirty-five years searching for beautiful books and manuscripts. I have a less exalted task in this paper—a quest for a less imperfect test. Ever since that fateful day in September 1930 when I began my teaching career at New York University until the present day when I am preparing tests for my future teachers of language arts, I have been on a continuous quest for more valid, more reliable, more objective, and more economical tests (in teacher time). The quest—like that of the Holy Grail—is far from ended but my forty-five year odyssey may be of interest to my fellow English teachers at all levels everywhere where tests are necessary. Millions of dollars are spent each year on standardized tests in reading, computational skills, and various aspects of the language arts. Additional millions are expended on Scholastic Aptitude tests, College Boards, professional tests in the sciences, law, and Civil Service. In short, testing in America is big, big business.

Yet the honest maker of tests will be the first to admit that the perfect test, like the perfect cure for cancer, or the perfect method of teaching, is still far beyond our grasp. At best we can only reach for an elusive goal, knowing well that *that* goal is far, far away.

I think that one of the problems is the failure to understand exactly what a test can do and what it cannot.

Avoiding the Dishonest in Testing

One of the most perceptive statements about the right place of tests in English (confessedly the only discipline about which I am even partly qualified to write) is found in Loban, Ryan and

---

Dr. Mersand is Chairman Emeritus of English at Jamaica High School and formerly Associate Professor of Education at York College.

[1]William Dana Orcutt, *In Quest of the Perfect Book*. Boston, Little, Brown, 1926.

Squire's *Teaching Language and Literature, Grades 7-12:*

> *"Tests and measurements concentrate on specific, well-defined skills. They have their niche, but what can be measured is by no means the same as what is important to teach. Paper and pencil testing has a highly limited scope. It serves to measure knowledge of spelling, and punctuation but fails to cope with power of the spoken word, appreciation of literature, and appreciation of written style. In comparison with other kinds of evidence, paper and pencil tests must be firmly assigned a relatively unimportant position."*[1]

Not only must we guard against expecting too much from pen and pencil tests, but we must never place ourselves in the categories of dishonest testmakers so brilliantly described by Orville Palmer in his article, "Seven Classic Ways of Grading Dishonestly." The former editor of the College Board programs of the Educational Testing Service, Princeton, New Jersey, knows whereof he speaks when he places these less than professional teachers in the following categories. Teachers test dishonestly:

1. by abdication (putting together crude and inadequate tests at the last minute and tailoring the course to ready-made tests)
2. by using the "carrots and clubs" system (grading on extraneous factors—deportment, penmanship, covers for reports, extra credit for reading)
3. by default (maintaining that tests are meaningless and giving so few that students' grades depend on an overly small sample of their performance)
4. by becoming zealots (grading everything except classroom posture)
5. by changing the rules in mid-game (shifting grading standards up or down to threaten or encourage)

[1]Walter Loban, Margaret Ryan, James R. Squire, *Teaching Language and Literature, Grades 7-12.* 2nd Edition. Harcourt Brace Jovanovich, 1969, p. 209.

6. by being a psychic grader (considering that tests are superfluous and inferior to intuition)
7. by anchoring the grading system "in the rainbow of an impossible perfection" (no one gets an A, few B's, and many fail)[3]

Understanding the Place of Testing

If the honest teacher wishes to avoid these pitfalls and strive to perfect his testing procedures, the first step, it is obvious, is an understanding of the place of testing in the total evaluation program. Although now almost a quarter of a century old, the distinction made by the Commission on the English Curriculum in *The English Langauge Arts* is still valid today. A comprehensive evaluation program attempts to achieve the following:

1. Identifies the needs of the learner
2. Provides important information for use in guidance
3. Stimulates self-direction
4. Aids in planning and replanning the language arts curriculum
5. Provides a record of growth and development for learners, instructors, and parents[4]

As will be seen, such aims cannot be achieved by dashing off a twenty-item true-false test in a few moments between class sessions. The intricacies involved in evaluating the outcomes of instruction in English can be seen as the teacher studies such a classic volume as Dora V. Smith's *Evaluating Instruction in Secondary School English* in which she reported her findings in New York State. Many of the principles and methods employed in this survey have significance today and will continue to be meaningful for a long time.[5]

Another excellent statement of criteria of a good evaluation

[3]Orville Palmer, "Seven Classic Ways of Grading Dishonestly." *English Journal*, Vol. 51, October 1962, pp. 464-467.

[4]Commission of the English Curriculum of the National Council of Teachers of English, *The English Language Arts*. New York, Appleton-Century-Crofts, 1952, pp. 417-440.

[5]Dora V. Smith, Evaluating Instruction in English. Urbana, Illinois, N.C.T.E., 1941.

program is found in DeBoer, Kaulfers, and Miller in their *Teaching Secondary English:*

1. Does the evaluation program include measurement of the extent to which the offerings in English are compatible with the needs and abilities of the students?
2. Does the evaluation program include adequate means for individual diagnosis of pupils' needs and difficulties?
3. Does the evaluation program provide for a cumulative record that can be passed on from teacher to teacher as the student progresses through school?
4. Does the evaluation program include tests of special abilities, such as spelling, vocabulary, or usage as diagnostic instruments to assist in locating possible causes of difficulties in using language for purposes of communication?
5. Does the evaluation program include appraisal of outcomes in the way of reading interests, speech habits (including usage outside of the English class), insights into human relations through literature, and other traits or qualifications cited as objectives of the course of study?
6. Does the program stress pupil participation in self-evaluation?
7. Does the evaluation program de-emphasize grades as ends in themselves?
8. Is the evaluation program appropriately differentiated to accommodate terminal as well as college-preparatory students?[6]

Desirable Characteristics of Good Testing

Proceeding from the total evaluation program specifically to the testing component, what do we hope to accomplish? Theodore W. Hipple in his recent *Teaching English in Secondary Schools* lists the following desirable characteristics:

1. It should provide students taking it an opportunity to learn from the test.

[6]John J. DeBoer, Walter F. Kaulfers, Helen Rand Miller, *Teaching Secondary English.* New York, McGraw-Hill, 1950, pp. 405-408.

2. It should be reliably scored.
3. It should permit all students to have at least some chance for success.
4. It should reflect broad coverage of whatver it is supposed to test.
5. It should be brief enough to be completed by most of the students in the time allotted for it.
6. It should be announced early enough to permit review and study by the student.
7. It should reflect student interest.[7]

Using these guidelines how does the teacher—especially the newly appointed one—go about preparing such tests, be they short answer or essays?

Preparing a Classroom Test

Soon after the beginning or experienced teacher starts the new term, he/she will be thinking of tests. In order to know where the class as a whole and each student stands as far as skill and knowledge are concerned, he/she will plan a diagnostic test or series of them. The summer vacation is a good time for forgetting much that has been learned in the preceding year. Students may be coming from lower schools in the same district or they may have transferred from another area. Mobility is a constant factor in education today.

Diagnostic Tests and What They Reveal

What can be learned by these diagnostic tests? Reading skills, spelling skills, grammatical knowledge can be diagnosed by objective, short-answer tests. Skill in written composition can be estimated to some extent by the first composition of the term. Other aspects of the English program such as appreciation of literature, attitudes toward languages, etc. are more difficult to perceive in a diagnostic test.

Later in the term the teacher will be interested in ascertaining

[7]Theodore W. Hipple, *Teaching English in Secondary Schools*. New York, The Macmillan Company, 1973, pp. 289-290.

the *achievement* of her students, and here again tests will be used. The end of the term is frequently another occasion for either a composite test (the Final Examination) or a series of shorter tests. The information obtained through these tests can be very valuable if the tests are properly constructed. They may indicate growth in language power. They may lead the teacher to change his/her methods. They may lead to individual counseling of students when progress or lack of it is indicated. Above all, a test should not be just a means of getting a grade for purposes of promotion or retention in the same grade.

Since the classroom test is still the most frequently used measurement of evaluation, we must understand what it is and what it can best do. As Sherman N. Tinkelman states in his helpful brochure, *Improving the Classroom Tests:*

> *"Before a single test question is written, therefore, the teacher should have clearly in mind what purpose the test is to serve, what skills and content areas are to be measured and what their relative weights should be. To ignore these factors at the outset in preparing a test is to court the danger that the test, if it constitutes a poor sample of questions, may provide a misleading picture of pupil progress."*[8]

Robert B. Carruthers, Supervisor of English of New York State, has written an excellent pamphlet on *Building Better English Tests* for the National Council of Teachers of English in which he reiterates time and time again the need for great care in preparing questions. Having been in charge of the State-Wide Regents Examinations in English for almost three decades, he speaks from long and arduous experience.[9]

Types of Tests

Basically tests today are of two types: the essay and the objective type (short answer). Much controversy has taken place

[8]Sherman N. Tinkelman, *Improving the Classroom Test.* Albany, New York State Education Department, 1957, and later revisions.

[9]Robert B. Carruthers, *Building Better English Tests.* Urbana, N.C.T.E., 1963.

for the past fifty years about the relative merits of each type. For many years the essay type was popular in our secondary schools. Then came the great interest in measurement and in testing in the 1920's. Objective-type tests were prepared and marketed, and whole libraries were written defending or attacking these short-answer tests. In the 1960's and 1970's there was renewed interest in the essay test, particularly as a guage of accomplishment in written English. The controversy will continue because of several factors, not the least of which is the time consumed in grading essay questions and the subjective element always present in the marketing.

Essay Questions

Of the values of the essay question, Tinkelman says:

> *"The fundamental contribution of the essay question is that it requires the pupil to develop an answer from his own background and fund of experience without benefit of suggested possibilities or alternatives, and to express that answer in his own words. By requiring pupils to present evidence, to evaluate, to analyze, and to solve problems or approach problems in a new way, essay questions can serve to measure some higher level abilities and thus contribute to knowledge about the pupil."*[10]

Examples of essay questions which were actually used in the English Comprehensive Examination in New York State follow:

1. The very nature of man makes it impossible for him to live "by bread alone;" he must live by faith—faith in an ideal or principle, faith in his country, faith in mankind. From the poems and essays you have read, choose any *four* selections. In *each* case show by definite references that the author expresses one of the kinds of faith mentioned. Give titles and authors.

[10]Tinkelman, *op. cit.*

2. "History records changes. Literature makes them real for us." Using two novels or a novel and a short story, explain how each of the two selections has made some phase of history more real to you. Refer directly to the incidents and portrayals in the stories to support your discussion. Give titles and authors.
3. From books that we read we see many pictures of family life, some of which reveal sympathy, understanding and cooperation among the members, while others reveal discord and lack of harmony. From one novel and one play you have read, select a family group (parents with children; children with one another; family with servants, guests, or relatives) and discuss what leads to cooperation or prevents cooperation. Give titles and authors.[11]

The major difficulty with the essay question is that the rating is subjective. In spite of almost forty years of attempting to make rating of test essays more objective, "considerable variation of assigned marks has been found, not only between one teacher and another, but in ratings by the same teacher from day to day." Among the interesting experiments in trying to determine more objective criteria is the study in the Spring, 1958 issue of the *Newsletter* of the Michigan Council of Teachers of English, "Evaluating a Theme."

Another helpful booklet, distributed by the State Education Department of New York is "Suggestions on the Ratings of Regents Examination Papers in English," with a special section on the rating of literature essay questions of the type described above.[12] Many local and state courses of study have indicated ways of correcting compositions. "The Guide for Evaluating Written Work" and "Criteria for Evaluating a Theme" from

[11]Regents Examinations in New York State are statewide examinations given at either the end of the third year or the fourth year (for English), and have been prepared for almost three-quarters of a century. Copies of these examinations are printed at the end of several handbooks commercially prepared by English teachers and/or supervisors. The handbooks are designed to help students review three or four years of English and to prepare them for the comprehensive examination given at the end of the semester or year.

[12]*Guideposts to English Language Arts*. Seattle, Bureau of Curriculum and Instruction, 1962, and later revisions.

Guideposts to the English Language Arts of Seattle Public Schools merit attention.[13]

Objective-Type Tests

Objective tests are usually of two types. In the first, the student *supplies* the answer; in the second, the student *selects* the answer from several supplied.

Eminent Victorians was written by _____.
Iago is a character in _____.
Wessex is the setting for the novels by _____.

The *advantages* of such questions are:

1. The answers are definite and specfic.
2. They can be rated quickly and accurately.
3. Wide sampling is possible, since a large number of questions can be answered in a unit of testing time.

The *disadvantages* are:

1. The teacher tends to concentrate on details.
2. You must be certain that there is only *one* correct answer to each question. Whereas the average student may give the answer you expected, the bright student who has read widely may think of several answers.

How to Write Good Completion Items

1. Write your item so that only one correct answer is possible. For example, in these items several answers are possible:
   Shakespeare wrote _____.
   (Plays, poems, etc.)
   Coleridge was born in _____. (This might be answered by the year, century, city, country.)
2. It is better to place the blank at the end of the statement, rather than at the beginning.

[13]Copies may be obtained from the Bureau of Testing and Measurement, State Education Department, Albany, New York.

3. Avoid extraneous clues to the correct answer.
   If the class has been studying forms of poetry and has learned about the *sonnet, ode, triolet,* you have given extraneous clues to this question.
   A long, serious poem is called an _____.
   (Obviously the student would not write *a sonnet* or *a triolet.*)

The Selection Type of Question

There are three forms: true-false, multiple choice, and matching.

True-False

In this question, the student writes *T* or *F* or *True* or *False,* depending upon the answer.

Ernest Hemingway won the Nobel Prize.
Shakespeare was born in the 17th century.
The Scarlet Letter was written in the early 19th century.

Advantages are:

1. These are fairly easy to write.
2. A large sampling can be covered in any given test period.
3. Scoring is objective and quick.

Disadvantages are:

1. It is limited almost entirely to testing facts on specific information.
2. Many of the larger outcomes of English instruction, such as generalizations, explanations, evaluations, cannot be tested by true-false questions.
3. There is encouragement to guess at the correct answer.
4. It is difficult to write questions that can be marked true or false without qualifications. Even a simple question like "Shakespeare wrote *Hamlet.*" may draw a negative answer from the bright student who has read something

about the Baconian Theory or who remembers that on the title page of his edition is *The Tragedy of Hamlet.*

How to Write Good True-False Items

1. Make statements that are absolutely true or false without qualifications or exceptions.
2. Avoid statements that contain elements that are partly true and partly false.
3. Avoid trick questions. (Students resent them.) Marjorie Kinnan Rawlings wrote *The Yearlin.* (The correct title is *The Yearling.*)
4. Avoid long involved statements with qualifying clauses.

Multiple-Choice Questions

These questions are very popular with the various College Boards, Civil Service, Army Intelligence and similar tests because the answers can be indicated on a separate sheet which can be scored electronically and quickly.

Advantages are:

1. The correct answer is limited to the choices given, thus eliminating any possibility of students' concern over universal application.
2. The guessing factor is less than in the true-false question.
3. Reasoning ability, fundamental understandings, as well as facts and principles may be tested.
4. Many items can be tested in a period of class time.
5. The scoring is objective and rapid.

Disadvantages are:

1. It is applicable when there are a number of related possible answers. Where the choices are obviously unrelated, the student can guess the answers easily.
2. It is time-consuming to prepare four or five choices for a single question.

Matching Items

A series of items is listed in a column to be matched with those in another column. It is advisable to have more possibilities than correct answers in the second column so that students who know some of the answers will not guess at the others by elimination.

Criteria of a Good Test

1. *Validity.* A well constructed test should measure a student's progress toward realization of the objectives of the course or unit.
2. *Reliability.* It should distinguish those who have mastered the subject from those who have not. It should consist of items carefully selected from the whole range of factual materials or principles taught. Poor administration of tests decreases reliability.
3. *Clarity.* It should be clear as to directions, construction of test items, assignment of credit, and should be free from ambiguity or error.
4. *Objectivity.* In evaluating an answer to a question, several teachers marking the same paper should arive at approximately the same mark.
5. *Administrability.* In a good test the exercises are arranged as to difficulty, convenience in scoring, and with provision for a reasonable allotment of time.
6. *Originality.* The test should challenge thought, stimulate imagination, arouse interest, provide for variety, and generally serve as a teaching as well as a testing device.

Last Words on Testing Have Not Been Said

Testing has been a subject for discussion, heated argument, and objective debate almost from the very beginning. The fact that so much time was spent at the 1974 convention of The National Council of English Teachers on discussion of testing proves that the final words have not yet been said, even though

every textbook on methodology in English has a substantial number of pages on the subjects of testing and evaluation. Perhaps the results coming out of the National Assessment of Education may provide us with clues for even better types of tests than those indicated on previous pages. At any rate, tests obviously are here to stay and it is the responsibility of every teacher of English to think long and seriously when preparing tests designed to indicate growth and/or achievement in all areas of the English Language Arts.

Although the search for the perfect test has been for me almost half a century long (far longer than it took Odysseus to return home from his wanderings), it has been a challenging one all along the way. Not Tennyson's Three Knights in *Gareth and Lynette,* but many more had to be met, conquered and the quest renewed. Will there ever be a perfect test? May I answer with: Will there ever be a perfect teacher or a perfect method? But, like Avis we must keep trying harder.